数理解析研究所講究録 1241 巻 2001 年 103-108

サイジング効果付き記憶制限準ニュートン法

静岡大学工学部　根岸達彦 (Negishi Tatsuhiko)　Shizuoka University
静岡大学工学部　八巻直一 (Yamaki Naokazu)　Shizuoka University
東京理科大学理学部　矢部　博 (Yabe Hiroshi)　Science University of Tokyo

1　概要

非線形関数の最小化のための数値解法として, 最急降下法の「大域的収束性」とニュートン法の「局所的に速い収束性」というそれぞれの長所をあわせ持つ準ニュートン法 (quasi-Newton method) がよく知られている. 準ニュートン法は無制約問題に対する数値解法の中では, 現在もっとも有力な方法の1つである. 近年, 準ニュートン法を大規模な問題に適用できるような工夫として, 記憶制限準ニュートン法が注目されている. 本研究の目的は, 記憶制限準ニュートン法に対して, サイジングが自然に適用される更新公式をつくることである. このために, 最適解までに速く収束させるために, 近似行列 H_k が満たす重要な条件であるセカント条件を拡張する. セカント条件を拡張することによって, 最適解までの収束を早めるサイジングと記憶領域を低減する記憶制限準ニュートン法を同時に成り立たせる.

2　非線形計画問題

対象とする非線形計画問題は以下の問題である.

$$\min f(x)$$

ただし,$x \in R^n$,$f : R^n \to R$ である.

本研究では, このような非線形計画問題の内,x の次元が大きい, 大規模な問題を考える.

3　非線形計画問題の解法

上で述べた非線形計画問題を反復法で解く. 反復法とは適当な初期点 x_0 から出発して, x_0 から出発して, $x_{k+1} = x_k + d_k$ によって点列 x_k を生成, 最終的に最適解 x^* に収束させるものである. 反復法で探索方向 d_k を求める解法として, 最急降下法, ニュートン法, 準ニュートン法がある. 探索方向 d_k は最急降下法では,

$$d_k = -\nabla f(x_k)$$

である. 最急降下法は目的関数の1階微分の情報しか用いていないから収束は遅いが, 探索方向が常に降下方向になる. ニュートン法では, 以下の連立1次方程式

$$\nabla^2 f(x_k) d_k = -\nabla f(x_k)$$

を解くことによって探索方向を求める. ニュートン法は目的関数の 2 階微分の情報を用いているので収束は速いが, 探索方向が必ずしも降下方向にならない. 準ニュートン法では, 以下の連立 1 次方程式

$$d_k = -H_k \nabla f(x_k)$$

を解くことによって探索方向を求める. ここで,H_k は目的関数のヘッセ行列の逆行列を近似する行列である. ニュートン法に近い速度を持っていて, 探索方向が常に降下方向になる.

4 準ニュートン法

準ニュートン法のアルゴリズムは, 以下のとおりである.

- **STEP0)** 初期点 x_1 を与え, $H_1 = \mathrm{I}$ とする. $k = 1$ とおく.
- **STEP1)** 探索方向 d_k を, 次のように求める.

$$d_k = -H_k g_k, \quad (g_k = \nabla f(x_k))$$

- **STEP2)** 停止判定をする.
- **STEP3)** ステップ幅 α_k を決定する.
- **STEP4)** x を次のように更新する.

$$x_{k+1} = x_k + \alpha_k d_k$$

- **STEP5)** H_k を更新し, $k = k + 1$ として STEP1 へ

記憶制限準ニュートン法は, 上のアルゴリズムにおいて, H_k を行列として保持せず, ベクトル演算によって求めるものである. 近似行列 H_k は

$$H_k \approx \nabla^2 f(x)^{-1}$$

のように近似し, 更新公式で計算していく. 以下に代表的な BFGS 公式を挙げる.

$$H_k = w_k[H_k - \frac{H_k y_k y_k^{\mathrm{T}} H_k}{y_k^{\mathrm{T}} H_k y_k} + (y_k^{\mathrm{T}} H_k y_k) u_k u_k^{\mathrm{T}}] + \frac{s_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}$$
$$u_k = \frac{H_k y_k}{y_k^{\mathrm{T}} H_k y_k} - \frac{s_k}{s_k s_k^{\mathrm{T}} y_k}$$

ここで,

$$s_k = x_{k+1} - x_k, \quad y_k = \nabla f(x_{k+1}) - \nabla f(x_{k+1})$$

5 セカント条件の拡張

近似行列 H_k が満たす重要な条件がセカント条件である．$\nabla^2 f(x_k)$ の近似行列を B_k，その逆行列を H_k としたときに，$\nabla f(x)$ のテイラー展開

$$\nabla f(x_k) = \nabla f(x_{k+1}) + \nabla^2 f(x_{k+1})(x_k - x_{k+1}) + \cdots$$

を $(x_k - x_{k+1})$ の項で打ち切ると $\nabla^2 f(x_{k+1})s_k = y_k$ なので，近似行列 B_{k+1} が

$$B_{k+1}s_k = y_k$$

を満たすことが要請される．これは s_k 方向で，$\nabla f(x_{k+1})$ を近似することを課すもので，セカント条件と呼ばれている．H 公式ではセカント条件は,

$$s_k = H_{k+1}y_k$$

となる. ここでは，セカント条件を拡張して,

$$\begin{aligned} S_{k+1} &= [s_1, s_2, \cdots, s_k], \\ Y_{k+1} &= [y_1, y_2, \cdots, y_k] \end{aligned}$$

と定め,

$$S_{k+1} = H_{k+1}Y_{k+1} \tag{1}$$

とする．記憶制限準ニュートン法は，過去 t ステップの情報を記憶する．すなわち,

$$s_{k-t}, s_{k-t+1}, \cdots, s_{k-1}$$

および

$$y_{k-t}, y_{k-t+1}, \cdots, y_{k-1}$$

を保持する．したがって，過去 t ステップに対して拡張されたセカント条件を考慮することが出来る.

6 拡張セカント条件を満たす公式

Yamaki and Yabe [3],Yabe and Yamaki [4] は，近似行列 H_{k+1} の生成において，拡張セカント条件を満たす公式を提案している. 公式は以下のとおりである.

$$\begin{aligned} &H_k = P_k + R_k \\ &v_k = P_k y_k - \frac{y_k^{\mathrm{T}} P_k y_k}{y_k^{\mathrm{T}} u_k} u_k \\ &u_k = s_k - R_k y_k \\ &R_k = S_k (Y_k^{\mathrm{T}} S_k)^{-1} S_k^{\mathrm{T}} \\ &P_{k+1} = w_k [P_k - \frac{P_k y_k y_k^{\mathrm{T}} P_k}{y_k^{\mathrm{T}} P_k y_k} + \frac{1}{y_k^{\mathrm{T}} P_k y_k} v_k v_k^{\mathrm{T}}] \\ &P_1 = \mathrm{I}, \quad R_1 = 0 \end{aligned} \tag{2}$$

式 (1) を満たす H_i を用いて，直線探索を行わない ($\alpha_k = 1$) 準ニュートン法によって生成された点列を $\{x_i\}$ とする．このとき，最小化される関数は２次関数

$$f(x) = \frac{1}{2}x^{\mathrm{T}}Qx - b, \tag{3}$$

ただし，Q は正定値対称，とすると，以下の性質がある．[3]

s が $s_1, \cdots, s_{k-1}$ と一次従属であり，H_k が正則であれば，$\nabla f(x_{k+1}) = 0$ である．一方，$s_1, \cdots, s_n$ が互いに一次独立であれば，$H_{n+1} = Q^{-1}$，$x_{n+2} = Q^{-1}b$ である．したがって，式 (1) を満たす更新公式を用いることによって，２次関数に対し，点列 $\{x_i\}$ は高々$n+1$ ステップで最適解に収束する．

7 サイジング

Oren [1] は，目的関数を２次モデルと仮定した場合, 近似行列 H_k の固有値を目的関数の最適解 x_* におけるヘッセ，行列 $Q = \nabla^2 f(x_*)$ の逆行列の固有値に単調に近づける工夫を提案している．このことは，

$$\tilde{H}_k = Q^{\frac{1}{2}} H_k Q^{-\frac{1}{2}}$$

とすると，次のスペクトル条件数 κ_{k+1} を単調に減少させることと同値である．

$$\kappa_k = \| \tilde{H}_k \| \cdot \| \tilde{H}_k^{-1} \|, \quad \kappa_k > \kappa_{k+1}$$

サイジングとは近似行列 H_k をそのまま更新するのではなく, 適当な正の数 w_k を H_k にかけてから更新することである. w_k は $w_k H_k$ の固有値の分布がヘッセ行列 $\nabla^2 f(x_k)$ の固有値の分布に近づくように選ばれる. このことによって, 計算効率を高めることが期待できる. Yabe and Yamaki[4] では，(2) に対する w_k の選び方として,

$$\begin{aligned}
w_1 &= \frac{(1-\psi_1)s_1^{\mathrm{T}}y_1}{y_1^{\mathrm{T}}H_1y_1} + \frac{\psi_1 s_1^{\mathrm{T}}g_1}{g_1^{\mathrm{T}}H_1y_1}, \quad \psi_1 \in [0,1] \\
w_k &= \frac{\psi_k^1 s_k^{\mathrm{T}}y_k}{y_k^{\mathrm{T}}H_ky_k} + \frac{\psi_k^2 s_k^{\mathrm{T}}g_k}{g_k^{\mathrm{T}}H_ky_k} + \psi_k^3 \\
&\sum_{i=1}^{3}\psi_k^i = 1, \quad \psi_k^1, \psi_k^2, \psi_k^3 \geq 0, \quad \text{for} \quad k \geq 2
\end{aligned}$$

を提案している．したがって，$k=1$ の場合には $H_1 = \mathrm{I}$ なので，簡単な計算によって w_1 が得られ，$k>1$ の場合は $w_k = 1$ でよいことがわかる．また，H_k は正定値対称であるとして，s_k は S_k の列ベクトルと一次独立であり，$w_k > 0, \phi_k \in [0,1]$ とする．そのとき，$w_k \tilde{H}_k$ の最小と最大の固有値の間に 1 が存在するなら，$\kappa(\tilde{H}_k) \geq \kappa(\tilde{H}_{k+1})$ である．

8 記憶制限付き準ニュートン法

Nocedal[2] は，通常の準ニュートン法では, H_k を保存するのに大量の記憶領域を必要とするのに対して, H_k を保存しないで探索方向 d_k を数本のベクトルの積形式で直接計算することによって，記憶領域を低減できることを示した．これを記憶制限準ニュートン法という．Nocedal と同様の変形をおこなうと，式 (2) は次のような積形式で表すことができる．

$$P_k = (\mathrm{I} - \frac{y_{k-1}u_{k-1}^{\mathrm{T}}}{y_{k-1}^{\mathrm{T}}u_{k-1}})^{\mathrm{T}} P_{k-1} (\mathrm{I} - \frac{y_{k-1}u_{k-1}^{\mathrm{T}}}{y_{k-1}^{\mathrm{T}}u_{k-1}})$$

$$Z_{k-1} = (\mathrm{I} - \frac{y_{k-1}u_{k-1}^{\mathrm{T}}}{y_{k-1}^{\mathrm{T}}u_{k-1}})$$

とおくと，P_k は次のようにあらわされる．

$$\begin{aligned} P_k &= w_1 Z_{k-1}^{\mathrm{T}} P_{k-1} Z_{k-1} \\ &= w_1 Z_{k-1}^{\mathrm{T}} \cdots Z_1^{\mathrm{T}} P_1 Z_1 \cdots Z_{k-1} \end{aligned}$$

このとき Z_{k-1} から Z_1 まで用いて計算するのではなく，s, u, y を t 個保存して,Z_{k-1} から Z_{k-t} まで用いて計算すれば，メモリは大幅に縮減できる．すなわち，$P_{k-t} = \mathrm{I}$ とし，

$$P_k = w_1 Z_{k-1}^{\mathrm{T}} \cdots Z_{k-t}^{\mathrm{T}} Z_{k-t} \cdots Z_{k-1}$$

のようにする．探索方向 d_k は

$$\begin{aligned} d_k &= -H_k g_k \\ &= -(P_k + R_k) g_k \\ &= -w_1 Z_{k-1}^{\mathrm{T}} \cdots Z_{k-t}^{\mathrm{T}} Z_{k-t} \cdots Z_{k-1} g_k - R_k g_k \end{aligned}$$

である．

上の式において，右辺の各項は g_k を右から掛けることからはじめると，順次 $t \times t$ の行列ないし，n 次元ベクトルが生成されるだけで計算できる．したがって，n が大きい場合，準ニュートン法が $n \times n$ の行列を記憶しなければならないのに比べて，格段に記憶容量が節減できる．

9　数値実験

以下の問題に対して数値実験を行った．

問題 1

$$f(x) = \sum_{i=1}^{n-1} (x_i - x_{i+1})^4 + \sum_{i=1}^{n} (x_i - 1)^2$$

問題 2

$$f(x) = \sum_{i=1}^{n-1} (x_i - x_{i+1})^2 + \sum_{i=1}^{n} (x_i - 1)^2$$

上記のどちらの問題も初期値を $x^{\mathrm{T}} = (-1, \cdots, -1, 0)$，収束条件を $\| f(x_k) \|_2 < 10^{-5}$，直線探索は *armijo* ルールを用い，記憶ステップは 5 として実行した．また，どちらの問題も 最適解は $x^{\mathrm{T}} = (1, \cdots, 1)$ である．

表 1　問題 1 の実験結果　最適解までのステップ数

次元	初期サイジング無し	初期サイジング有り	BFGS 公式
50	23	14	18
1000	31	17	
2000	76	23	

表 2　問題 2 の実験結果　最適解までのステップ数

次元	初期サイジング無し	初期サイジング有り	BFGS 公式
50	23	17	20
1000	23	18	
2000	22	18	

表 1，表 2 で，初期サイジング無しは我々の公式で初期サイジングを行わなかったものを意味し，初期サイジング有りは初期サイジングを行ったものを意味する．BFGS は初期サイジング無しの BFGS 公式で計算したもので，表中の空白は計算不能になったものである．

参考文献

[1] Oren,S.S.,Self-scaling variable metric(SSVM) algorithms, Part2; Management Science,20,863-874(1974)

[2] Nocedal,J,Updating quasi-Newton matrices with limited storage, Mathematics of Computation,35.773-782(1980)

[3] Yamaki,N and Yabe,H, A family of the quasi-Newton methods, TRU Mathematics,16-1,49-54(1980)

[4] Yabe,H and Yamaki,N, Some properties of Oren's SSVM type algorithm for unconstrained minimization, TRU Mathematics 16-2,103-111(1980)

[5] 矢部　博，八巻　直一，非線形計画法，朝倉書店，(1999)